

# SEJF

## Słownik elektroniczny jednostek frazeologicznych Electronic Dictionary of Multi-Word Units

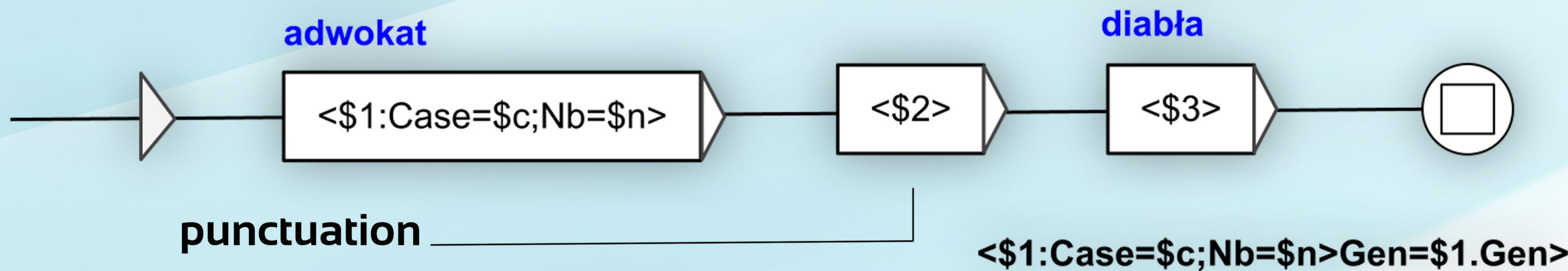
about 5000 entries of Polish multi-word expressions (nouns, adjectives, adverbs)  
available here: <http://zil.ipipan.waw.pl/SEJF> under the Creative Commons BY-SA license

Monika Czerepowicka, University of Warmia and Mazury in Olsztyn - lexicography  
Agata Savary, Université François Rabelais de Tours - automatic inflection and validation

### Description of multi-word units - a multistage procedure

#### 1. segmentation

strings of letters (words) and their separators (space, punctuation marks) are treated equally: expression *adwokat diabła* ("devil's advocate") consists of three segments: 1. *adwokat*; 2. space; 3. *diabła*. Each of them receives an appropriate interpretation.



#### 2. morphological description

word (a sequence of letters between text separators) is interpreted as a form of a specific lemma with assigned morphological characteristics (tag). Morphological analyzer (Morfeusz SGJP) suggests the basic form for each word as well as all interpretive possibilities and detects every homonymy, cf:

\$	Czlon	Lemat	Tag	Odmieniny
1	krewni	krewny	subst:pl:nom:m1	<input checked="" type="checkbox"/>
2			sp	<input type="checkbox"/>
3	i	i	conj	<input type="checkbox"/>
4			sp	<input type="checkbox"/>
5	znajomi	znajomy	subst:pl:nom:m1	<input checked="" type="checkbox"/>
6			sp	<input type="checkbox"/>
7	królika	królik	subst:pl:nom:m2	<input type="checkbox"/>

Wybierz prawidłowy tag  
krewny;adj:pl:nom:m1:pos  
krewny;adj:pl:voc:m1:pos  
krewny;adj:pl:nom:p1:pos  
krewny;adj:pl:voc:p1:pos  
krewny;subst:pl:nom:m1  
krewny;subst:pl:voc:m1

#### 3. formal description by means of graphs

In the graph we recode the parameters which are actually inflected. Cf :

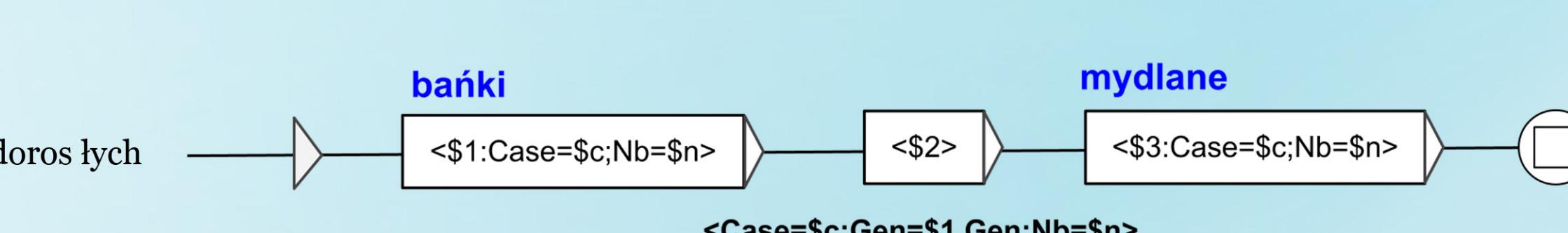


### Grammatical and lexicographical problems:

#### lemmatisation

- Z kolei " Bańki [bańka:subst:pl:nom:f] mydlane [mydlany:adj:pl:nom:n:pos]" " to wiersze dla dorosłych  
party będzie polegala na wszelkobecnych  
w czołówce tabeli prysło jak  
bańka [bańka:subst:sg:nom:f] mydlana [mydlany:adj:sg:f:pos]

Widzial ktoś coś podobnego  
. - Przyjechaliśmy po



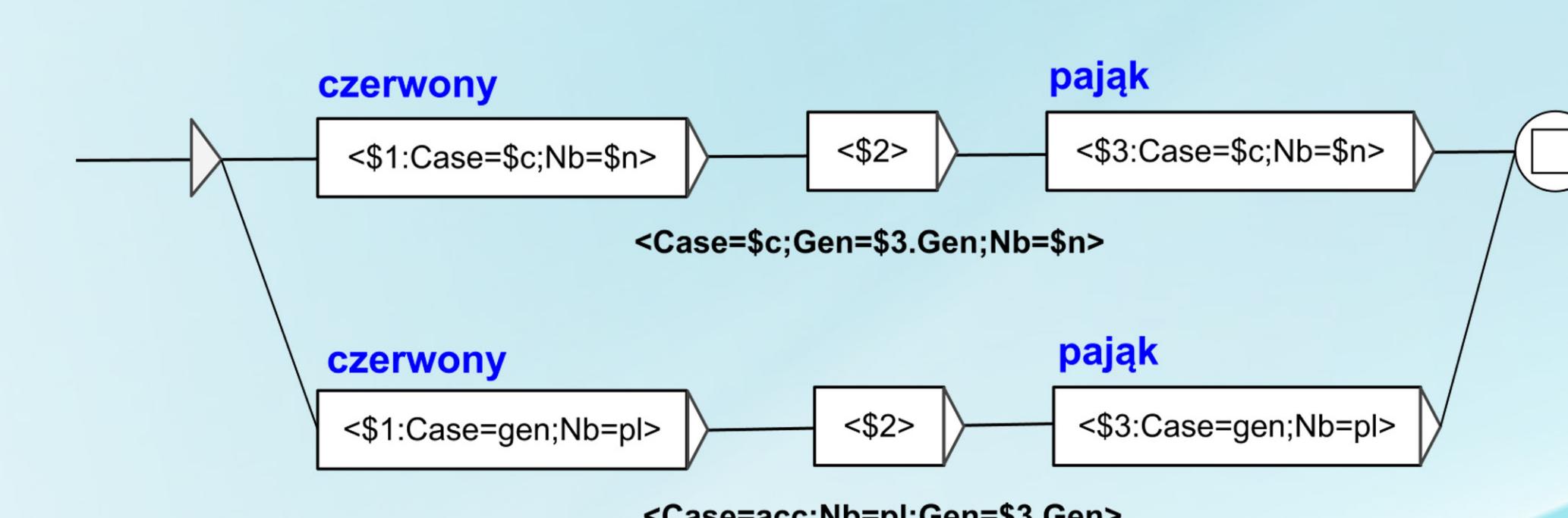
bańki mydlane @ bańki mydlane subst:sg:nom:f  
bańki mydlane @ bańki mydlane subst:sg:gen:f  
bańki mydlanych bańki mydlane subst:sg:gen:f  
bańki mydlany @ bańki mydlane subst:pl:gen:f  
bańki mydlany @ bańki mydlane subst:pl:acc:f  
bańki mydlane @ bańki mydlane subst:sg:inst:f  
bańki mydlany @ bańki mydlane subst:pl:loc:f  
bańki mydlany @ bańki mydlane subst:sg:vo:f  
bańki mydlane @ bańki mydlane subst:pl:voc:f

#### fluctuation of the grammar gender

ustroju narzuconego części Europej przesz  
Towarzysze z CK. Zdechli  
że głosowalem na tych  
w opuszczeniu mieszkaniu czekała na  
Po pięciu dzwonkach włączyła się  
o tej porze zgłasza się  
więc dzwoniliam na jego

czerwone [czerwony:adj:pl:acc:m2:pos] pajaki  
[pajak:subst:pl:acc:m2]  
czerwony [czerwony:adj:sg:nom:m2:pos] pajak  
[pajak:subst:sg:nom:m2]  
czerwonych [czerwony:adj:pl:gen:m2:pos] pajaków  
[pajak:subst:pl:gen:m2:pos]

z Brukseli. Na szczęście  
. Grzebią...  
.-

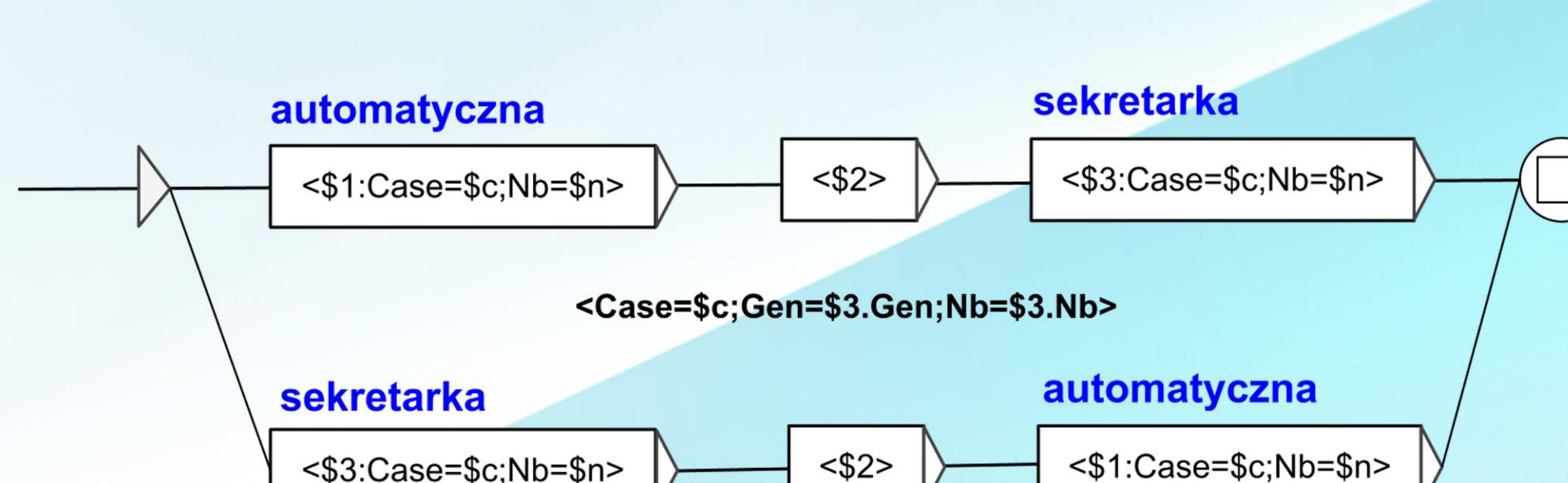


czerwony pajak @ czerwony pajak subst:sg:nom:m2  
czerwone pajak @ czerwony pajak subst:pl:nom:f  
czerwonego pajak @ czerwony pajak subst:sg:gen:m2  
czerwonych pajaków @ czerwony pajak subst:pl:gen:f  
czerwonym pajak @ czerwony pajak subst:sg:inst:m2  
czerwonego pajaka @ czerwony pajak subst:sg:loc:f  
czerwonym pajakiem @ czerwony pajak subst:pl:loc:f  
czerwonych pajakach @ czerwony pajak subst:sg:voc:m2  
czerwony pajak @ czerwony pajak subst:sg:voc:f  
czerwone pajaki @ czerwony pajak subst:pl:voc:f

#### variable word order

w opuszczeniu mieszkaniu czekała na  
automatycznej [automatyczny:adj:sg:loc:f:pos] wiadomość od niej  
sekretarce [sekretarka:subst:sg:loc:f]  
automatyczna [automatyczny:adj:sg:nom:f:pos] . Usłyszałem swój głos  
sekretarka [sekretarka:subst:sg:acc:f]  
sekretarka [sekretarka:subst:sg:nom:f] . Po godz. 7  
automatyczna [automatyczny:adj:sg:nom:f:pos]  
sekretarkę [sekretarka:subst:sg:acc:f]  
automatyczna [automatyczny:adj:sg:acc:f:pos]

zwiastunie od niej  
. Usłyszałem swój głos  
. Po godz. 7  
10 razy dziennie, żeby



automatyczna sekretarka @ automatyczna sekretarka subst:sg:nom:f  
automatyczne sekretarki @ automatyczna sekretarka subst:sg:nom:f  
automatycznej sekretarki @ automatyczna sekretarka subst:sg:gen:f  
automatycznej sekretarki @ automatyczna sekretarka subst:sg:inst:f  
automatycznej sekretarki @ automatyczna sekretarka subst:sg:loc:f  
automatycznej sekretarki @ automatyczna sekretarka subst:sg:voc:f  
automatyczne sekretarki @ automatyczna sekretarka subst:sg:voc:f  
automatyczna sekretarka subst:sg:acc:f  
automatyczna sekretarka subst:sg:gen:f  
automatyczna sekretarka subst:sg:inst:f  
automatyczna sekretarka subst:sg:loc:f  
automatyczna sekretarka subst:sg:voc:f  
automatyczna sekretarka subst:sg:acc:f